Using a Growth Mindset Intervention to Help Ninth-Graders

An Independent Evaluation of the National Study of Learning Mindsets

Pei Zhu
Ivonne Garcia
Kate Boxer
Sidhant Wadhera
Erick Alonzo

mdrc
BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

Using a Growth Mindset Intervention to Help Ninth-Graders

An Independent Evaluation of the National Study of Learning Mindsets

Pei Zhu
Ivonne Garcia
Kate Boxer
Sidhant Wadhera
Erick Alonzo

November 2019

mdrc
BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

The independent evaluation of the National Study of Learning Mindsets is funded by the Bill & Melinda Gates Foundation.

Dissemination of MDRC publications is supported by the following organizations and individuals that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, Arnold Ventures, Charles and Lynn Schusterman Family Foundation, The Edna McConnell Clark Foundation, Ford Foundation, The George Gund Foundation, Daniel and Corinne Goldman, The Harry and Jeanette Weinberg Foundation, Inc., The JPB Foundation, The Joyce Foundation, The Kresge Foundation, and Sandler Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

Overview

The transition from middle school to high school can be challenging for adolescents as they are faced with new academic challenges and an unfamiliar social environment. Students who successfully navigate this transition and pass their ninth-grade classes are far more likely to graduate from high school with their peers and attend college than those who fail courses in the ninth grade. The growing awareness of the importance of the first year of high school for future success has prompted the development of interventions for ninth-graders.

One type of such intervention uses psychological tools to communicate to young people that their brains can grow "stronger." These positive beliefs about intelligence, often referred to as "growth mindset" beliefs, are expected to result in academic resilience, which can lead to better academic performance.

To test whether a growth mindset intervention could improve the academic performance of adolescents, the National Study of Learning Mindsets (NSLM) implemented a low-cost growth mindset intervention specifically designed for ninth-graders in a nationally representative sample of regular high schools during the 2015-2016 school year. The national study used a student-level randomized controlled trial design to gauge the impacts of this intervention on students' mindsets about intelligence, their own behaviors, and their academic achievements. With support from the Bill & Melinda Gates Foundation, MDRC reviewed the data from the NSLM and conducted an independent evaluation of this growth mindset intervention.

Key Findings

This evaluation found the following:

- The intervention changed students' self-reported mindset beliefs, their attitudes toward efforts and failure, and their views on academic challenges.
- Immediately after the intervention, students were more likely to take on challenging academic tasks.
- The intervention produced statistically significant impacts on students' average academic performance, improving their average grade point average (GPA) as well as their math GPA, and reducing the proportion of students with failing grades.
- Certain groups of students and schools might benefit more from the intervention than others. These groups include students with relatively low academic achievement before the intervention, schools in the midrange of the academic performance spectrum, and schools where students are more inclined to take on challenging tasks.

These findings are substantively consistent with the results published by the NSLM research team.

Contents

List of Exhibits

Figure

Box

Table

Acknowledgments

This report and the study upon which it is based are funded by the Bill & Melinda Gates Foundation. The authors would like to thank Brad Bernatek and Andrew Sokatch from the Foundation, and Fred Doolittle and Michael Weiss from MDRC for their guidance and comments on this work.

The study benefited greatly from support and cooperation from important members of the research team behind the National Study of Learning Mindsets: David Yeager, Chandra Muller, and William R. Reynolds from the University of Texas, Austin, and Paul Hanselman from the University of California, Irvine responded promptly and patiently to our many information requests. Furthermore, we received many helpful comments from Lisa Quay and Shanette Porter from the Mindset Scholars Network.

Finally, this study would not have been possible without the support from our MDRC colleagues: Kate Gualtieri, Hannah Power, and Alpesh Shah provided fiscal oversight. Margaret Bald, Ali Tufel, Anaga Dalal, and Joshua Malbin reviewed earlier drafts of the report, provided valuable comments, and supported the editing of the report. Ann Kottner prepared the report for publication.

The Authors

Introduction

Although graduation rates have improved in recent years, too many students still do not complete high school. In the United States, over half a million adolescents fail to graduate high school with their peers each year.[1] Without a high school diploma, these young people will face more challenges in future education and career paths in an increasingly complex job market, and are at higher risk of poverty.[2]

The transition from middle school to high school can be challenging for students. Many stray off the graduation path when they first enter high school. Ninth-graders must adjust to more demanding course work, develop relationships with new teachers and peers, and respond to unprecedented academic expectations and social pressures. As a result, they may feel less confident about their abilities as learners and struggle to overcome academic challenges. Research shows that ninth-graders on average have the lowest grade point averages (GPAs) and the most unexcused absences and misbehavior referrals compared with students in all other high school grade levels.[3] Many who fall behind in ninth grade have a harder time recovering credits and face a greater risk of dropping out. The University of Chicago Consortium for School Research and others have shown that students who fail their ninth-grade classes are far less likely to graduate from high school and attend college.[4] The growing awareness of the importance of the first year of high school for future success has prompted the development of interventions targeting ninth-graders.

One type of intervention is designed to improve students' academic success by changing their attitudes and behaviors in relation to school and schoolwork. These interventions do not provide students with academic instructions. Instead, they aim to alter their view about their potential, their sense of belonging in a new environment, and their perception of challenges. Rigorous evaluations have shown that this kind of intervention can have an impact on students' academic outcomes.[5] The growth mindset intervention evaluated here falls into this category.

Figure 1 illustrates how changing mindset beliefs can lead to changes in academic outcomes. The growth mindset intervention uses online modules and exercises to convey the message that individuals can change their intellectual ability by exerting effort, trying different strategies, and seeking help. This shift from a "fixed mindset," which holds one's intelligence cannot be changed, to a "growth mindset," which holds the contrary, is expected to motivate students to take on more academic challenges and to overcome difficulties they encounter at school. These changes in mindsets and behaviors are expected to improve their academic achievement.

---

[1]McFarland, Stark, and Cui (2016).

[2]Goldin and Katz (2008); Hanushek and Woessmann (2008).

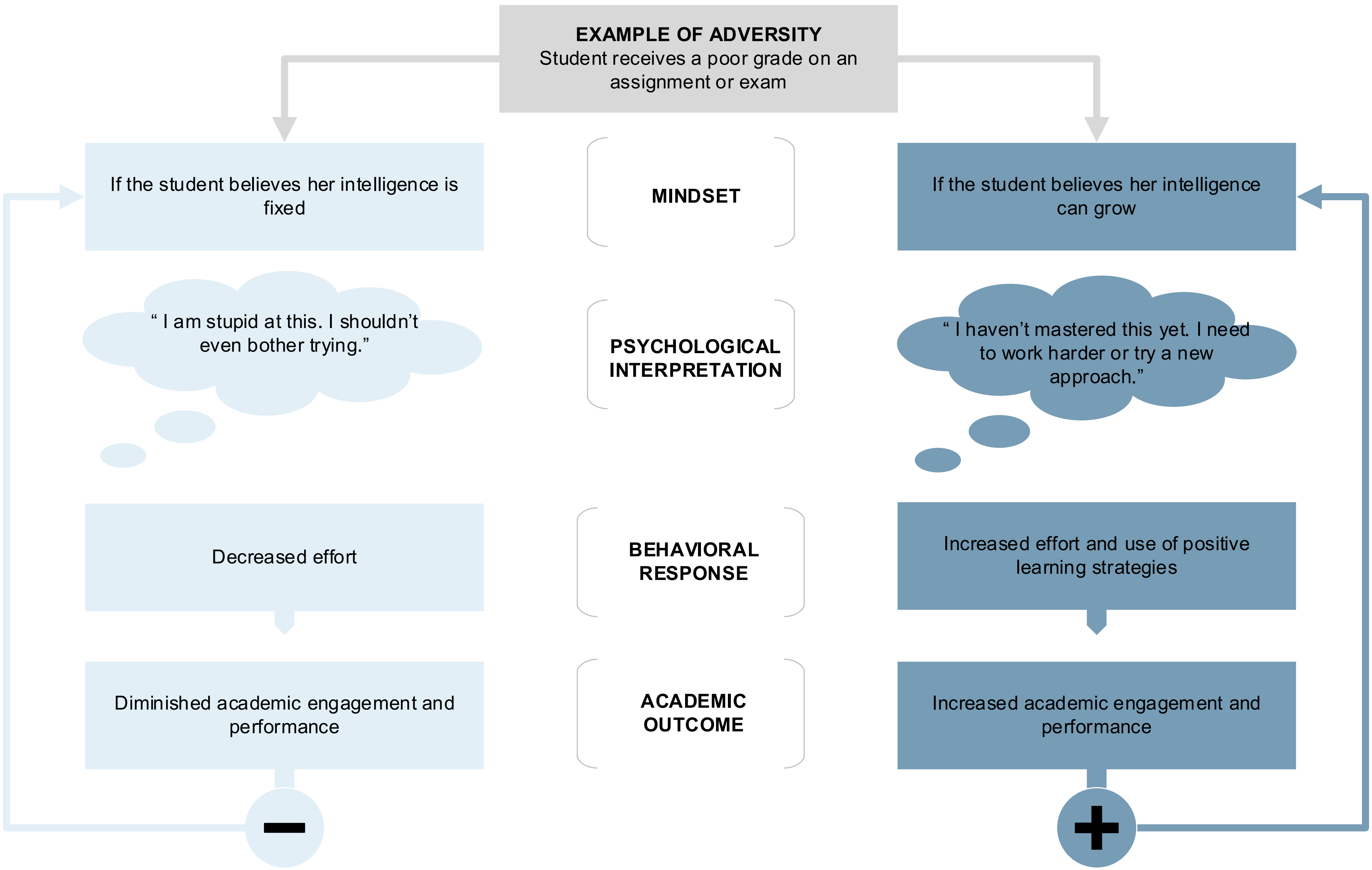
[3]McCallumore and Sparapani (2010).

[4]Allensworth and Easton (2005).

[5]See Yeager and Walton (2011) for a review.

Figure 1

How a Growth Mindset Intervention Is Expected to Affect Student Outcomes



SOURCE: Copied with permission from the Mindset Scholars Network (October 2019)

The specific version of the growth mindset intervention under evaluation here was developed, implemented, and studied by an interdisciplinary team of psychologists, sociologists, education researchers, statisticians, and economists at major universities around the United States (collectively referred to as the National Study of Learning Mindsets [NSLM] research team), with the support of the Mindset Scholars Network and the Center for Advanced Study in the Behavioral Sciences.[6] This version was adapted from previous growth mindset interventions to address the specific challenges that occur in the transition to high school, such as changes in self-perceptions of academic abilities and increased academic rigor.[7] The NSLM implemented this intervention in a nationally representative sample of regular high schools in a student-level randomized controlled trial during school year 2015-2016. Box 1 provides more information about the national study.

In 2018, the Mindset Scholars Network invited MDRC to review the data collected for the national study and to conduct an independent evaluation of the growth mindset intervention. Specifically, MDRC reviewed the data provided by the NSLM research team and verified the data-processing method. The MDRC team then conducted independent data analyses following the research approach outlined in the preregistered analysis plan for the NSLM study.[8] The MDRC analyses focused on the following questions proposed in the plan:

- What was the average impact of a growth mindset intervention on the mindset beliefs, challenge-seeking behaviors, and academic achievement of ninth-grade students in regular U.S. public high schools?
- How did the impact of a growth mindset intervention on ninth-graders' academic achievement vary among students?
- How did the impact of a growth mindset intervention on ninth-graders' academic achievement vary among schools?

The rest of this report presents the findings for each of the research questions. The MDRC team also documented the data-processing procedures used to create the analytic data file for its reevaluation of the study.[9] A restricted-use data file will be made available to the research community and will provide an opportunity for other researchers to make further use of this data set.

[6]Mindset Scholars Network (2015b).

[7]Aronson, Fried, and Good (2002); Good, Aronson, and Inzlicht (2003); Blackwell, Trzesniewski, and Dweck (2007); Paunesku et al. (2015); Yeager et al. (2016).

[8]The preregistered analysis plan for the NSLM can be found at https://osf.io/tn6g4.

[9]The construction of the key variables used in the MDRC analyses is summarized in Appendix A.

Box 1

The National Study of Learning Mindsets

A Growth Mindset Intervention Tailored for Ninth-Graders*

The intervention consists of two 25-minute, self-administered online modules designed to communicate the message that the brain can grow "stronger" in response to efforts such as trying new strategies and seeking help from experts. This stronger brain can help students achieve meaningful goals. Students were also asked to internalize the message by teaching it to a future struggling ninth-grader. All materials were written for the vocabulary, conceptual sophistication, and interests of adolescents entering high school and use arguments that might be most relevant or persuasive for 14- to 15-year-olds.

Ninth-Graders from a Nationally Representative Sample of High Schools†

The NSLM research team worked with a third-party data-collection and research firm — ICF International — to select a sample of 139 high schools through stratified random sampling from a national high school population of 11,221 regular public high schools that met a certain set of criteria. Among them, 76 schools agreed to participate in the study, and 65 of these 76 schools provided records data to ICF. In schools with 300 or fewer ninth-graders, all students were included in the sample. In schools with more than 300 ninth-graders, the study randomly sampled a set of required core classes to ensure approximately 300 students (or about 10 classes) from the school would be included in the sample. Findings can be considered to represent ninth-graders in the national population of regular public high schools in the United States.

A Student-Level Randomized Controlled Trial

In the fall of 2015, sample students were asked to participate in online training on the topic of brain development. When students logged into a computer system, the system randomly assigned them with equal probability to either receive the intervention (the program group) or not (the control group). While the study was in progress, no person in the study knew what group any student belonged to. Students in the program group received the online modules about the growth mindset and were asked to answer reflective questions in a survey. Students in the control group read a brief article about the brain and were also asked to answer survey questions. Instead of learning about the brain's malleability, control group students learned about basic brain functions and the areas of the brain responsible for them. The two experimental conditions were designed to look very similar to prevent students and instructors from knowing which groups they were in and to discourage students from comparing their materials. Later in the school year, these study participants were asked to complete a second module designed the same way.

NOTES: *Yeager et al. (2016).

†Specifically, eligible schools were regular public high schools that were not charter schools, schools serving special populations, alternative schools, institutions of adult education, or schools run by the Department of Defense or the Bureau of Indian Affairs. High schools with grades lower than ninth or with fewer than 25 ninth-graders were also excluded. See Tipton, Yeager, Iachan, and Schneider (in press) for additional details about the sampling frame and the stratified probability sampling process used for school selection.

Impact Findings for All Students

The MDRC team used a sample that slightly deviates from that of the NSLM: Two schools were excluded because the course-grade information they provided was vague for course names and grading periods, and including this information would require additional assumptions for the data interpretation. The final sample used by the MDRC team consists of 11,888 ninth-graders from 63 high schools across the United States. This sample is limited to students with nonmissing GPA scores at the end of ninth grade. Among these students, 5,916 (49.76 percent) were randomly assigned to the program group, while 5,972 (50.24 percent) were randomly assigned to the control group.[10] This section presents findings on the impacts of the growth mindset intervention based on data from this sample of students.

- **The intervention changed students' self-reported beliefs about intelligence as intended.**

The growth mindset intervention was designed to change students' beliefs about the malleability of their intelligence. Several survey questions embedded at the end of the second online session captured different aspects of this belief. Responses to these questions are on a scale of 1 ("strongly disagree") to 6 ("strongly agree"), and higher values in the responses indicate stronger beliefs in a fixed mindset. Figure 2 shows that the program group students' responses to these questions differ from those of their control group counterparts.

For example, compared with the control group, the program group students were less likely to agree with statements such as "You have a certain amount of intelligence, and you really can't do much to change it," and "Your intelligence is something about you that you can't change very much." They were also less likely to hold the belief that "being a 'math person' or not is something that you really can't change. Some people are good at math and other people aren't." The differences in the responses to these questions between the two groups range from 0.22 to 0.36 standard deviations in effect size and are all statistically significant. The intervention also changed students' views on learning and schooling: The program group students were less likely to think that one of their main goals was to avoid looking "dumb" in front of their peers.

There is also some evidence that the intervention might have changed students' attitudes toward failure. Students were asked, about a hypothetical failure in math, whether "This means I am probably not very good at math" or whether "I can get a higher score next time if I find a better way to study."[11] Their responses to these two statements measure their beliefs in the view that failures only confirm their lack of ability. The last two rows in Figure 2 show that the intervention affected students' responses to the first statement but not to the second.

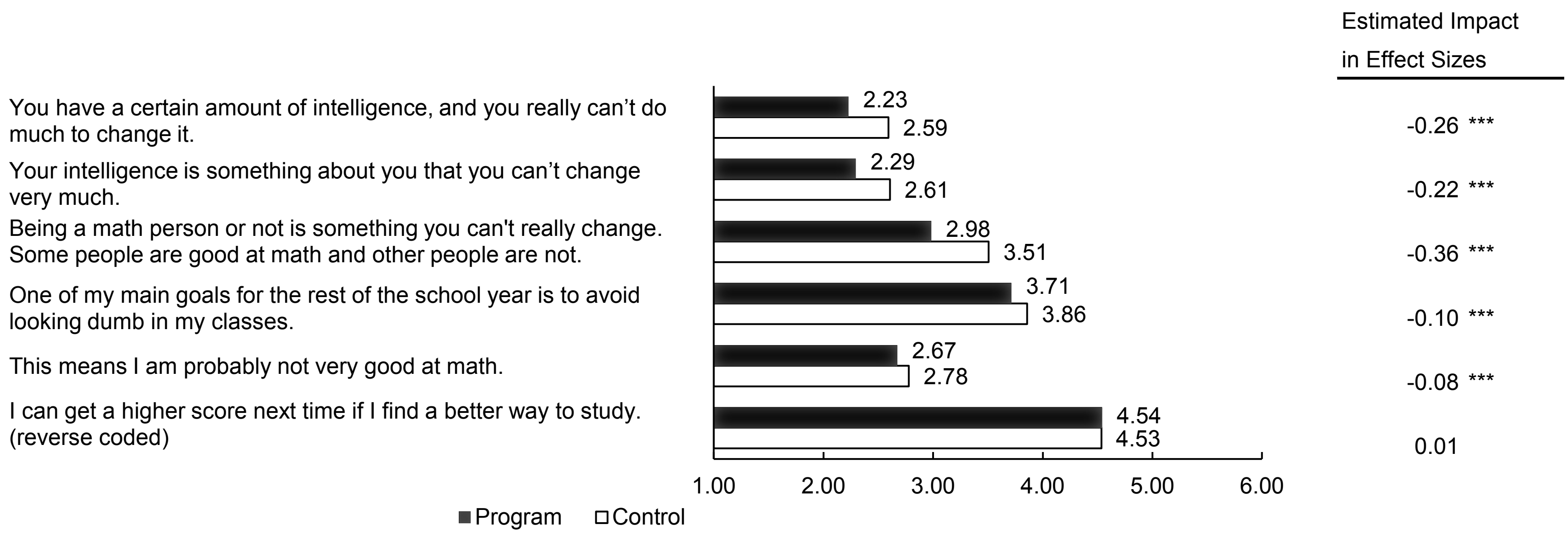
---

[10]Appendix B provides detailed information about the characteristics of this sample of schools and students.

[11]To make its values align with those of other outcome measures, the responses to this last question were "reverse coded," meaning that a higher value indicated stronger fixed mindset beliefs and vice versa.

Figure 2

The Intervention Changed Students' Attitudes and Beliefs



SOURCE: MDRC calculations based on student responses to survey questions embedded in the second online session.

NOTES: Individual student responses were presented on a scale from 1 to 6, ranging from "strongly disagree" (= 1) to "strongly agree" (= 6).
The sample includes ninth-grade students in the 63 sample schools for whom ninth-grade achievement information is available. The number of observations ranges from 10,642 to 10,690 due to different response rates across these variables.
A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

The intervention also changed students' challenge-seeking intentions. The survey asked students to choose between the following:

- An easy review that has math problems they already know how to solve, and they will probably get most of the answers right without having to think very much
- A hard challenge that has math problems they don't know how to solve, and they will probably get most of the problems wrong, but they might learn something new

Fifty-one percent of the program group students said they would choose the hard challenge, while only 38 percent of the control group students said so.[12] This difference is an indication that the program group students were more likely to believe that their ability can be changed by taking on challenging tasks.

Overall, these results show that the intervention pushed students' self-reported beliefs from a fixed mindset toward a growth mindset and changed their view of academic challenges as intended. Students' responses to these questions were collected right at the end of the second online session, so it is possible that they were giving socially desirable answers, and the observed differences in their responses might not reflect their true beliefs. Longer-term follow-up data could help assess this possibility.

- **Students were more likely to take on challenging academic tasks after the intervention.**

People with a growth mindset believe that working through challenges is a way to increase ability; therefore, they are more likely to pursue challenging academic material.[13] To test this hypothesis, the study administered a "Make a Math Worksheet" task to all students toward the end of the second online session. This task asked students to create a math worksheet by selecting math problems with different difficulty levels.[14] The choices include "easy" items that are probably below their ability level and they probably will not learn much from, "moderate" items that are suitable for their ability levels and they might learn a medium amount from, and "hard" items that may be challenging but from which they might learn a lot. The easy item carries a value of 1 for difficulty level, and the difficulty values for the moderate and hard items are 2 and 3, respectively.

Analyses of the number of items each student chose and their corresponding difficulty levels demonstrate that the intervention led students to take on more academic challenges. The program group students picked more hard items and fewer easy items than the control group

[12]The p-value is less than 0.001 for this estimated impact.

[13]Dweck and Leggett (1988) and Mueller and Dweck (1998).

[14]Students were told later in the module that they did not have time to complete the worksheet, so they did not have to solve the problems.

students did. On average, the program group students selected about 0.5 more hard items than the control group students; about 39 percent of the program group students chose more hard items than easy items, compared with about 31 percent of the control group students who did so. The average difficulty level of the items chosen by the program group students was about 0.1 point (on a 1 to 3 scale) higher than that of control group students' choices, and the total difficulty level across all items chosen by the program group students was 1.09 points higher than that of the control group students' choices. All of these differences are statistically significant.[15]

- **The intervention produced positive impacts on students' academic performances.**

Using school records, the team constructed three measures to capture students' postprogram academic performances.[16] These measures are as follows:

- **Average GPA**: This is a continuous measure of student grades. It ranges from 0 to 4.3 points and is calculated as the average grade across four core subject areas: math, English/language arts, science, and social studies. It summarizes a student's overall academic performance and is the key outcome of interest.

- **Poor-performance indicator**: This is a binary variable indicating whether a student has an average grade of 1.0 or lower. It serves as a proxy for whether a student is academically on track and is often used by schools as an early indicator of success in high school. This measure, especially when used for ninth-graders, is of policy relevance because many states have adopted it as a key metric for school accountability as a result of the Every Student Succeeds Act (ESSA).

- **Math GPA**: This is the grade for the core math course. Math is considered a key subject that could influence students' performance in other courses and their general academic performance.[17] There are also societal stereotypes about fixed math intelligence, which raises interest about whether this intervention could affect students' math performance.

In schools where the intervention occurred in the fall of 2015, these outcomes are calculated based on the average of the fall and spring semester grades; in schools where the intervention took place in the spring of 2016, they are calculated based on the spring semester grades only.

---

[15]Appendix Table C.1 presents details of these findings.

[16]Details about core course coding and the construction of these outcome measures are in Appendix A.

[17]Lee (2012).

Figure 3 demonstrates that the growth mindset intervention produced statistically significant effects on all three achievement measures.[18] Specifically, it produced a 0.05 point impact on students' average GPA (effect size = 0.04), increasing it from the 2.55 points that they would have scored in the absence of the intervention (the control group) to 2.59 points with the intervention (the program group). Relatedly, the intervention significantly reduced the number of students scoring GPAs of 1.0 or below from 32.1 percent to 29.7 percent, a decrease of 2.4 percentage points, representing a risk reduction of 7.5 percent.[19] The intervention also had a positive effect on students' math GPA, increasing it from 2.42 points to 2.48 points on average (effect size = 0.05).

Impact Variation Among Student Subgroups

Prior studies suggest that the growth mindset intervention could be particularly beneficial for some students. For example, studies have found that interventions targeting student growth mindset might be especially helpful for academic underperformers because these students might encounter more academic difficulties, and the growth mindset intervention could affect their interpretation of these difficulties and help them better cope with the challenges. These underperformers also might have a larger-than-average margin for improvement in their academic performances.[20]

Similarly, students who started out with stronger fixed mindset beliefs might be more likely to have their attitudes and beliefs affected by the intervention because they have a large margin for change, and such changes might lead to larger gains in their academic performance. However, it might also be harder for such a brief intervention to affect individuals' strong beliefs. Research has shown that the growth mindset program could have a differential benefit for other students who might suffer from stereotypes that cast them as not academically strong because of who they are or their family background. For example, students of color, girls, and students from families who live in poverty all could have their learning suppressed by negative stereotypes that claim they likely cannot succeed academically; girls in particular face negative stereotypes about their abilities in math.[21] The growth mindset beliefs might enable these students to overcome such social identity concerns, and they, in turn, might benefit from the

---

[18]All findings remain statistically significant at the 5 percent level with the Benjamini-Hochberg adjustment to account for testing multiple hypotheses (Hochberg and Benjamini, 1990).

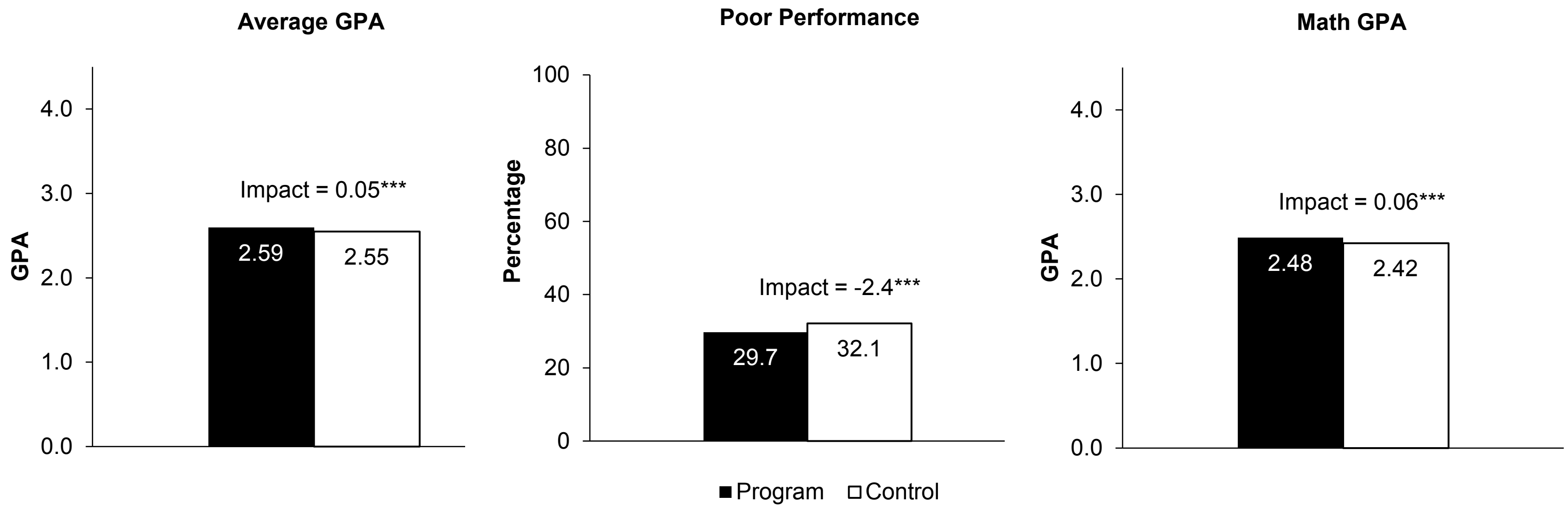
[19]The risk reduction rate is calculated as the ratio between reduction (2.4 percent) and counterfactual level (32.1 percent).

[20]Paunesku et al. (2015).

[21]Aronson, Fried, and Good (2002); Good, Aronson, and Inzlicht (2003); and Blackwell, Trzesniewksi, and Dweck (2007). There is some evidence that these students had stronger fixed mindset beliefs than their counterparts before the intervention: The preprogram fixed mindset measure (based on two survey questions, on a 1 to 6 scale, with higher values indicating stronger fixed mindset beliefs) was 2.72 for girls and 2.68 for boys, 2.80 for minority students and 2.57 for white students, and 2.86 for students eligible for free or reduced-price lunch and 2.53 for those who were not eligible.

Figure 3

The Intervention Increased Students' Average and Math GPAs and Reduced the Proportion of Students with Poor Performance (All Students)



SOURCES: Student responses to survey questions before the first online module and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average. The sample includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The number of observations is 11,888 for average GPA and the poor-performance indicator and 10,853 for math GPA.

Poor performance is indicated when average GPA is 1.0 or lower.

The GPA is measured on a scale of 0.0 to 4.3.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

intervention more than their peers who do not face negative stereotypes about their groups. This study explored these hypotheses by examining whether the program's effects on students' academic achievement vary across subgroups of students.

- **The program impacts varied by students' academic backgrounds: The program had an effect on academic performance for lower-performing students, but not for higher-performing ones.**

To explore the hypothesis that the growth mindset intervention might be especially beneficial for academic underperformers, the team defined a student as lower-performing if his or her preprogram average GPA was lower than the median GPA level in the school and as higher-performing if the preprogram GPA was higher than the school median.[22]

Figure 4 shows that the intervention produced effects on all three outcomes for the lower-performing students, increasing both their average GPA and their math GPA by 0.06 GPA points (effect sizes = 0.05) and reducing the proportion of students with poor performances by 3.3 percentage points. The impacts on the higher-performing group, in contrast, are virtually zero. The differences in impacts between these two subgroups are statistically significant for two of the three measures.

- **The impacts of the intervention did not differ across other subgroups defined by students' background characteristics or beliefs about intelligence when they joined the study.**

The program impacts on academic outcomes experienced by students starting out with stronger fixed mindset beliefs,[23] students of color, girls, and students eligible for free or reduced-price lunch (FRPL) were not different from those experienced by students starting out with weaker fixed mindset beliefs, white students, boys, or students not eligible for FRPL, respectively.[24] These results are somewhat surprising given that one might expect some of these groups to have had a larger margin for mindset changes than others, so they might also have been expected to benefit more from an intervention targeting fixed mindset beliefs. This no-difference finding is especially surprising for the subgroups defined by students' preprogram fixed mindset beliefs. In fact, the program impacts on students' postprogram fixed mindset beliefs and their challenge-seeking behavior vary significantly by students' initial mindset

---

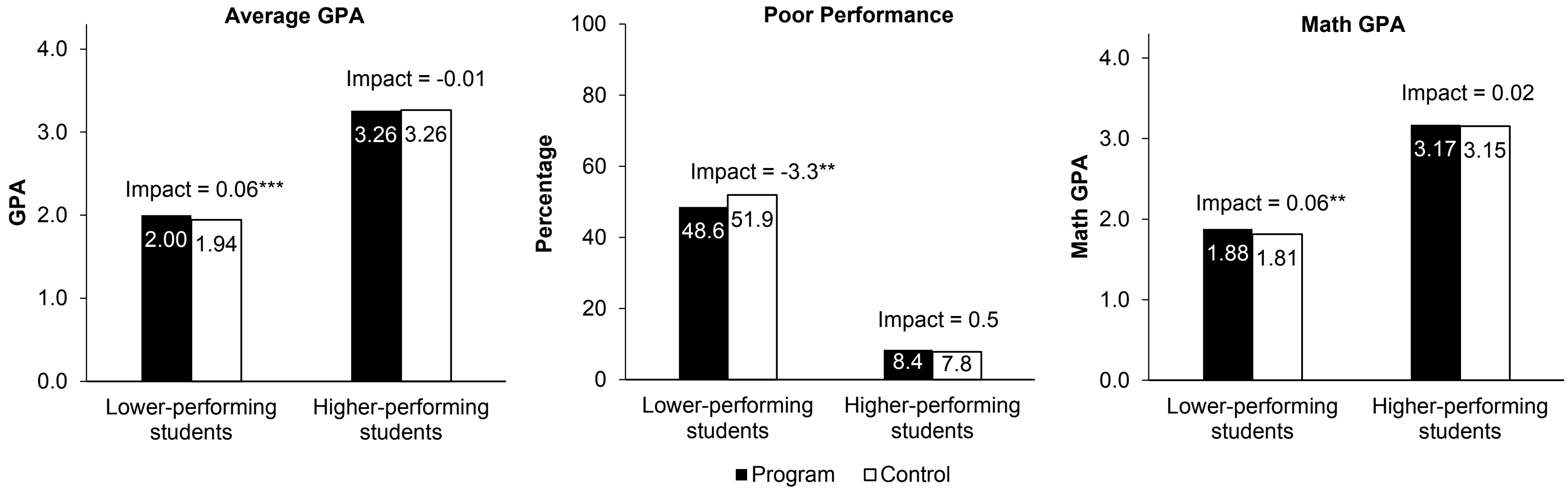
[22]Subgroups were defined based on students' preprogram GPAs. Students without a preprogram GPA were excluded from the analysis reported here. As a sensitivity test, the team used a profile analysis routine in MPlus to impute the group assignment for students with missing preprogram GPAs. Subgroup impact estimates based on the imputed subgroup designation are similar to those reported here (Panel A of Appendix Table C.2).

[23]Students are defined as having stronger fixed mindset beliefs if their self-reported fixed mindset ratings were higher than the school median rating.

[24]Appendix Table C.2 presents impact findings on academic outcomes for these student-level subgroups.

Figure 4

The Intervention Improved Academic Performance for Lower-Performing Students



SOURCES: Student responses to survey questions before the first online module and student record data from school years 2014-2015 and 2015-2016.

NOTES: The sample includes ninth-grade students in the 63 schools for whom pre- and postprogram achievement information is available. Students whose preprogram GPAs were below their school median GPAs are in the lower-performing subgroup, and those with preprogram GPAs above school medians are in the higher-performing subgroup.The number of observations for lower-performing students is 5,503 for average GPA and the poor-performance indicator and 4,992 for math GPA. The number of observations for higher-performing students is 5,274 for average GPA and the poor-performance indicator and 4,787 for math GPA.

The GPA is measured on a scale of 0.0 to 4.3.

Poor performance is indicated when average GPA is 1.0 or lower.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Two-tailed t-tests were used to test differences in estimated impacts between the two subgroups. The p-values for the differences in impacts between the two subgroups are 0.001 for the average GPA, 0.004 for the poor-performance indicator, and 0.320 for math GPA.

rating, but this variation is not observed for the program impact on student achievement.[25] In addition, students' initial fixed mindset belief rating is only weakly associated with their academic achievement level before the intervention, which means that students in the lower-performing subgroup are not necessarily those in the strong initial fixed mindset subgroup.[26] In fact, about 49 percent of lower-performing students reported weak fixed mindset ratings before the intervention. These factors might offer some clues to the pattern of findings reported here.

Impact Variation Among School Subgroups

Assessing of the relationship between program impact and various school characteristics can help answer the question "What types of schools might benefit the most from the growth mindset intervention?" This section follows the preregistered analysis plan and focuses on schools' prior achievement levels as a moderating factor for impact variation, but it also explores other moderating factors, such as the growth mindset climate in schools.

- **The impacts of the growth mindset intervention on academic outcomes are moderated by the school's overall prior academic achievement level.**

The preregistered analysis plan of the NSLM study hypothesized that schools with the lowest level of past academic achievements might not benefit much from the growth mindset intervention because they might not have adequate instructional resources to support student mindset change. Further, even if the program affected student mindset beliefs, it may be less consequential in schools with inadequate instruction or an unsafe environment. In contrast, schools in the middle range of the achievement distribution are expected to produce large effects, especially for previously lower-performing students, because these schools might already have adequate resources and a supportive environment, and they can benefit from an improvement in students' motivation that comes from the shift from fixed to growth mindset beliefs due to the intervention. The prediction for the high-performing schools is less clear: These schools could have a "ceiling effect" since they are already performing well academically and there is limited room for improvement. On the other hand, these schools could stand to benefit from a change in the mindset culture if they have all the right academic conditions and had a fixed mindset culture before the intervention.

The NSLM research team constructed the measure of school achievement level with data from multiple sources and divided schools in the targeted national population into three subgroups based on this measure: low-level schools whose achievement measure is in the bottom 25th percentile among all sample schools, medium-level schools that are in the 25th to

---

[25]One unit change in the preprogram fixed mindset rating is associated with a 0.07 point reduction in the program impact on the postprogram fixed mindset rating (p-value = 0.003), and is associated with a reduction of 0.08 (the number of hard questions chosen) in the program impact on challenge-seeking behavior (p-value = 0.059). The association between the preprogram fixed mindset rating and the program impact on GPA is not different from zero.

[26]The correlation between these two measures is -0.16.

75th percentile range, and high-level schools that are in the top 25th percentile.[27] Among the 63 schools used in the MDRC evaluation, 12 are in the low-performing category, 36 are in the medium group, and the remaining 15 are in the high-performing group.

Figure 5 shows that the schools with medium-level achievement experienced positive and significant impacts on students' average GPA, while the impacts on the schools with lower or higher achievement levels were much smaller in magnitude and are not statistically different from zero. This pattern holds for all students (Figure 5, top panel) and for the lower-performing students (Figure 5, bottom panel). Findings for the other two academic outcomes are presented in Appendix C. The difference in the impacts between the medium-level schools and other schools is statistically significant for two of the three academic outcomes for all students and is statistically significant for one of the outcomes for the lower-performing student group.[28] Overall, these results suggest that schools in the middle range of the prior achievement distribution could benefit more from the intervention than those with lower or higher achievement levels.

- **The impacts of the growth mindset intervention on academic outcomes vary by the prevalence of student challenge-seeking behavior in the school.**

Schools with a culture that is supportive of growth mindset beliefs and challenge-seeking behaviors could provide a favorable environment to sustain and enhance the behavioral changes induced by the growth mindset intervention and could lead to improved academic performances.

Following the preregistered analysis plan, this study uses two measures to capture schools' growth mindset environment. The first measure is a school average fixed mindset rating based on all students' responses to two relevant questions asked before they started viewing the training materials during the first online session.[29] This measure is subjective and is dependent on an individual's interpretation of a fixed mindset; therefore, it could potentially

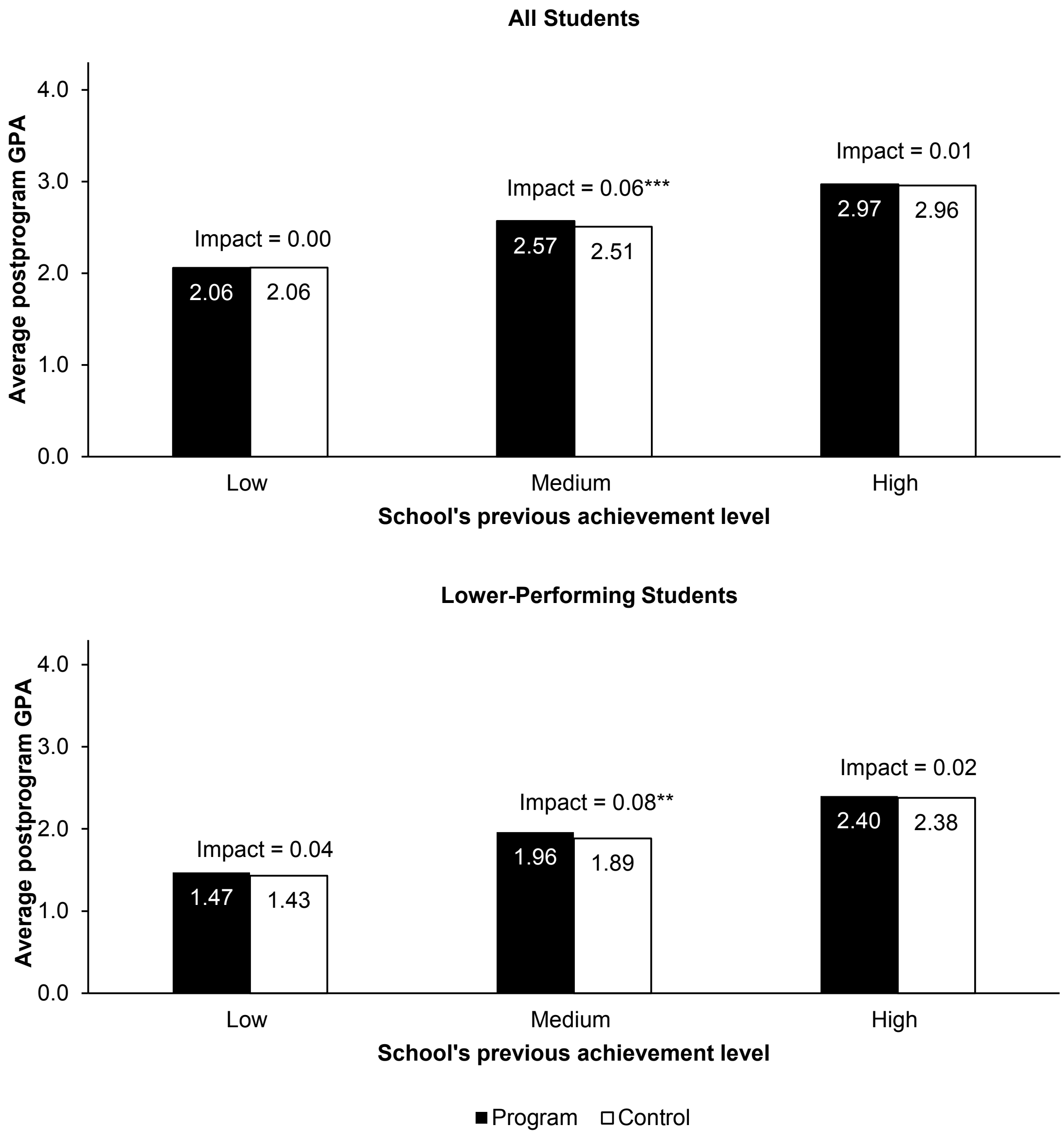
[27]These sources include the average state test scores for schools obtained from the Great Schools website, school average PSAT scores and Calculus AB and English (literature and language) AP participation rates obtained from the College Board, and state proficiency levels for math and reading for the eighth grade obtained from the National Assessment of Educational Progress (NAEP). See Tipton, Yeager, Iachan, and Schneider (in press) for more details.

[28]Details of these findings are in Appendix Table C.4. For all students, the p-values for the differences in impacts between these two groups are 0.111, 0.076, and 0.027 for average GPA, poor-performance indicator, and math GPA, respectively; for lower-performing students, the p-values are 0.278, 0.056, and 0.308 for the same set of outcomes.

[29]The two questions asked to what extent they agreed with the statements "You have a certain amount of intelligence, and you really can't do much to change it," and "Your intelligence is something about you that you can't change very much." Students' responses to these two questions from both the program and control groups are averaged across the questions and aggregated to the school level to capture the prevalence of fixed mindset thinking, or mindset environment, in a school.

Figure 5

Program Impact Varies by School's Previous Achievement Level, for All Students and for Lower-Performing Students



(continued)

Figure 5 (continued)

SOURCES: Student responses to survey questions before the first online session and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average, measured on a scale of 0.0 to 4.3. The sample for the top panel includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The sample sizes are 1,687, 6,287, and 3,914 for low, median, and high previous school achievement levels, respectively.

The sample for the bottom panel includes ninth-grade students in the 63 schools for whom both pre- and postprogram achievement information is available and whose preprogram average GPAs were below their school median GPAs.The sample sizes are 778, 2,883, and 1,842 for low, median, and high previous school achievement levels, respectively.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Two-tailed t-tests were used to test differences in estimated impacts between the medium group and the low and high groups combined. For all students, the p-value for the differences in impacts is 0.111. For lower-performing students, the p-value for the differences in impacts is 0.278.

suffer from reference bias.[30] The second measure uses the average number of challenging math problems chosen by the control group students for the "Make a Math Worksheet" task at the end of the second online session to measure the prevalence of challenge-seeking behavior in a school. It measures the mindset environment through "action" instead of "belief," and therefore it is more objective and less likely to suffer from reference bias. A school is categorized as having a "more/less supportive climate for growth mindset" if its average fixed mindset rating is below/above the median value across all participating schools or if its challenge-seeking behavior value is above/below the median.

Figure 6 shows that there is some evidence that schools with a more supportive environment for growth mindset might benefit more from the intervention.[31] When school environment was measured by the prevalence of challenge-seeking behavior in schools, the program produced a positive and statistically significant impact on average GPA (effect = 0.09 points, effect size = 0.08) for students in the more supportive schools, while the impact for the less supportive schools was virtually zero (Figure 6, top panel).

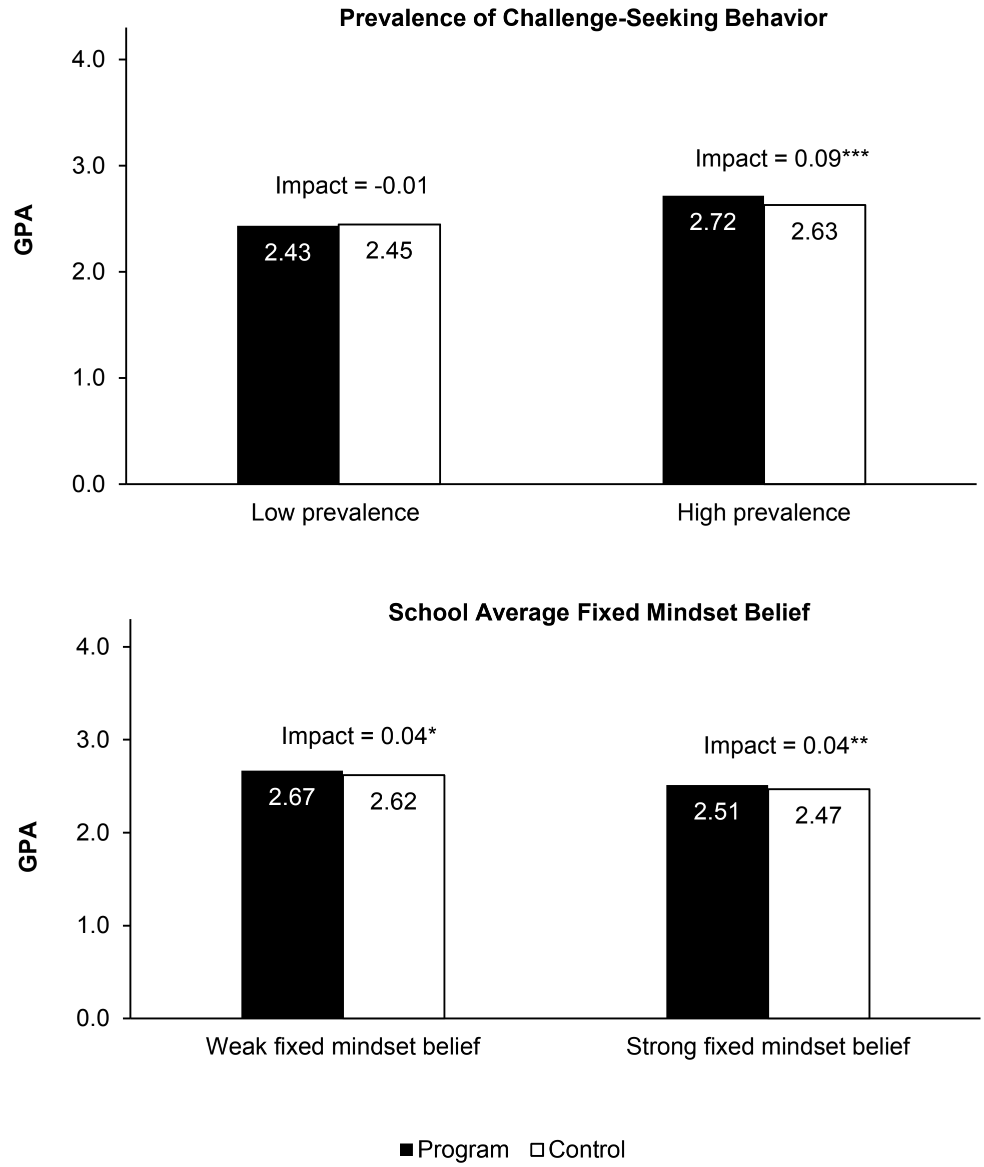
However, when the study schools were split into two subgroups based on their average self-reported fixed mindset beliefs, the impacts of the growth mindset intervention did not differ between the two subgroups: For all students, the estimated program impacts on the average GPA were 0.04 points (effect size = 0.04) for both subgroups, and the difference in impacts between these two groups was virtually zero (Figure 6, bottom panel). While there is no clear explanation for the different findings between these two sets of subgroup analyses, the correlation between students' challenge-seeking behavior and their fixed mindset belief ratings

[30]Duckworth and Yeager (2015). Reference bias here refers to the kind of distortion in responses that comes from students holding different standards (in this case, different understandings of fixed mindset beliefs) by which they make judgments.

[31]Appendix Table C.4 provides findings for all students, as well as for lower-performing students. The patterns of the findings are essentially the same for these two samples.

Figure 6

Program Impact Varies by the Prevalence of Challenge-Seeking Behavior in School, but Not by School Average Fixed Mindset Belief (All Students)



SOURCES: Student survey responses and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average, measured on a scale of 0.0 to 4.3. The sample for the top panel includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The sample sizes are 5,064 and 6,824 for the subgroups of schools with low and high prevalence of challenge-seeking behavior, respectively. The sample sizes are 7,133 and 4,755 for the subgroups of schools with weak and strong fixed mindset beliefs, respectively.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Two-tailed t-tests were used to test differences in estimated impacts between the two subgroups. The p-value for the difference in impact between the two subgroups defined by fixed mindset belief ratings is 0.956; the p-value for the difference in impact between the two subgroups defined by the prevalence of challenge-seeking behaviors is less than 0.001.

among the control group students is only -0.06 across students and is -0.46 across schools, suggesting that these two measures might be capturing different aspects of the school mindset environment.

Discussion

The impact findings show that the growth mindset intervention implemented by the NSLM study significantly changed students' self-reported attitudes about their mindset, moving them more toward the growth mindset beliefs and away from the fixed mindset beliefs. It also led them to be more likely to take on challenging academic tasks. Most important, the intervention improved the academic performance of ninth-graders in a nationally representative sample of public high schools: It increased the average GPA of a typical ninth-grader by around 0.05 points, or 0.04 standard deviations in effect size. The intervention also reduced the probability of failing core courses by about 2.4 percentage points for the average ninth-grader. These findings are substantively consistent with the findings reported by the NSLM research team.[32]

The magnitude of these impact findings is on par with the impacts on academic outcomes reported by other high school interventions. For example, 16 high school-level interventions supported by the federal Investing in Innovation (i3) fund report an average effect of 0.05 standard deviations in effect size on student academic outcomes.[33] This effect size of the impact on average GPA is also equivalent to about the 40th percentile in the distribution of effect sizes based on impacts on 481 academic outcomes reported by 242 randomized controlled trials of educational interventions across grade levels.[34]

It is important to consider the cost of the intervention when interpreting the magnitude of its impacts. The growth mindset intervention is brief and requires a total of less than an hour of training time from students. It is easily expanded to a larger scale, as demonstrated by the NSLM study. Most important, it can be delivered potentially at no material cost because the materials will be freely available and can be delivered through the internet, and it does not require professional development for school staff members. These attributes are in stark contrast with many successful educational interventions that are resource-intensive. For example, one study found that the median per-pupil cost of 68 educational interventions is $882.[35] The Enhanced Reading Opportunity Study, an evaluation of supplemental literacy programs targeting lower-performing ninth-grade students, found that the programs improved students' GPAs in core subject areas by an effect size of 0.07 standard deviations at a cost of $1,931 per student per year.[36]

---

[32]Yeager et al. (2019). Appendix D provides more detailed comparisons of the key findings from these two studies.

[33]This calculation is based on numbers reported in Appendix D of Boulay et al. (2018).

[34]Kraft (2018), Table 1.

[35]Kraft (2018), Table 1.

[36]Somers et al. (2010).

This study also identified, through subgroup analyses, certain types of students and schools whose academic performances might benefit most from the growth mindset intervention. Such groups include students with relatively low academic achievement before the intervention, schools in the midrange of the academic performance spectrum, and schools where students are more inclined to take on challenging tasks. This is useful information for policymakers and practitioners seeking to implement the intervention. More moderation and mediation analyses, although beyond the scope of the current evaluation, could provide insight on other mechanisms at work and could help better target the intervention and improve its effectiveness. In addition, while this evaluation was able to focus only on the immediate impacts of the intervention, it would be of interest to follow up with the sample students and collect their longer-term outcomes to see if the observed effects are sustainable over time.

Appendix A

Data Processing and the Construction of Key Measures

Working with the National Study of Learning Mindsets (NSLM) research team, the MDRC team processed the student-level transcript data collected by ICF International and constructed outcome measures based on these data. This appendix summarizes this process and provides brief descriptions of other data used in the current study. Details about the data-processing and measure-construction procedures can be found in the documents accompanying the restricted-use file created for this study. The restricted-use file will be deposited with the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan.

Processing Student Course Grade Data

The role the MDRC team played at the data-processing stage was to standardize the core course grades reported in each school's transcript data so that common measures of student academic performances could be constructed from these data and used for analysis purposes. The MDRC team carried out this task in two steps: First, the team identified course grades for each of the four core subject areas of interest for each grading period. These four areas are English/language arts, math, science, and social studies — the core subject areas that students need to cover to fulfill graduation and diploma requirements. Second, the team identified the pre- and postprogram period for each school based on the timing of the intervention. Results from these two steps were then used to construct the grade point average (GPA) for the period before and immediately after the intervention. The team conducted these processes without knowledge of students' program status and the potential impacts of various coding decisions. This process is summarized below.

Step 1: Identifying Courses in Core Subject Areas

The transcript data were provided to the study by many different districts with many different naming conventions for course titles. The MDRC team followed an iterative process to standardize the course data and identify courses for each of the core subject areas for each grading period by school.

To begin, to the extent possible, the MDRC team relied on course descriptions and course names in a school's course catalogs to identify core classes offered by each school. For schools with no available catalog, the team compiled a generic list of required courses for each subject area based on information from the following documents:

- National Forum on Education Statistics. 2011. *Prior-to-Secondary School Course Classification System: School Codes for the Exchange of Data (SCED)* (NFES 2011–801). Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Bradby, Denise, Rosio Pedroso, and Andy Rogers. 2007. *Secondary School Course Classification System: School Codes for the Exchange of Data (SCED)* (NCES 2007-341). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Since the team did not have access to any middle school course catalogs, a generic list was also generated for Grade 8 courses based on the above sources.

When a course name matched with the course catalog or the generic list, the team coded the course accordingly. When the course name did not match with courses in either source, the team made a judgment call by examining cross-tabulations of course enrollments in each grading period. For example, a course titled "Geo" could be geometry (math), geology (science), or geography (social studies). But if the cross-tabulations showed that "Geo" was mutually exclusive with Algebra 1 or 2 and often occurred with biology and world history, then the coders might infer that the course was geometry, not geology or geography. The MDRC team developed routines for flagging ambiguities so that these decisions were recorded and explained, and the course coding can be reproduced.

Next, this course-level file was merged with the student-level course file, and students with missing core courses were flagged. The team used two approaches to identify potential core courses for these students. First, the team examined the course-taking patterns of these students to see if they were enrolled in foundational courses in core subject areas. Taking these courses was considered as fulfilling graduation requirements even though these courses were usually not listed as core courses in the catalog. Hence, the team manually recoded these courses to corresponding core subject areas. Second, the team used text patterns to detect all courses that could potentially be core courses and manually recoded the course designations if two coders agreed on the decisions.

At the end of this step, the team created a student-level data file that contained course grades for each of the four core subject areas.

Step 2: Identifying Grading Period and Calculating Grades

One wrinkle in identifying the grading period was that some schools reported multiple grades for a given course in a marking period. This situation could occur, for example, if the school reported cumulative grades, quarter grades, and final grades; if the school allowed students to retake a course and reported both grades; or for other reasons. The team identified 10 schools with more than 5 percent of their students affected by the multiple-grade issue in either eighth grade or ninth grade. After consultation with the NSLM research team and having ICF make contact with some schools again for clarification, the MDRC team devised custom school-based rules that dictated how to determine which grade was a student's isolated semester grade for a given course, largely resolving this issue for 8 of the 10 identified schools. There were no clear solutions for the remaining two schools, and the team decided to exclude them from the analyses reported here.[1]

[1]For schools that have less than 5 percent students with multiple grades, the team used the average across the multiple grades as the final grade for a given course in a grading period for the affected cases.

Next, the team standardized all numeric and letter grades across all the schools to a scale of 0 to 4.3. The conversion made a priority of the letter grades and only used the numeric grades if the letter grades were not reported.

After the conversion, the team standardized grading periods across schools into four periods: eighth grade, ninth-grade first semester, ninth-grade second semester, and ninth grade. The team then averaged the grades across the four core subjects to calculate the average GPA in each grading period. Depending on when the intervention was given within a school, MDRC assigned different grading period GPAs as the pre- and postprogram academic outcomes for each student. Specifically:

- If the intervention was given in the fall in a school, then a student's preprogram GPA was the average (that is, across semesters) or final GPA in the eighth grade, and the postprogram GPA was the average GPA in the ninth grade.

- If the intervention was given at the beginning of the spring semester, then a student's preprogram GPA was the average ninth-grade first-semester GPA, or the average GPA across ninth-grade quarters or terms before the intervention was given in the spring, and the postprogram GPA was the average ninth-grade second-semester GPA.

This definition of preprogram and postprogram grading period is consistent with the preregistered analysis plan. In addition to the average GPA across all four core subject areas, the team also calculated preprogram and postprogram GPAs for each core subject area.

By the end of this process, the MDRC team had created a student-level file that contains pre- and postprogram GPAs, average and by subject area, for each student.

Data Used in the MDRC Evaluation and Their Sources

The NSLM research team provided MDRC access to all the data collected by ICF. The MDRC team worked with the NSLM research team to review and refine these data and focused on the following data elements for the analytic purpose of this report.

First, a set of variables describing student characteristics before the start of the intervention was used to assess the similarity of the program and control groups and to serve as control variable covariates in the impact estimation model. These variables were:

- Gender

- Race/ethnicity

- Individualized educational program (IEP) status

- Eligibility status for the free or reduced-price lunch (FRPL) program

- GPA before the random assignment
- Self-reported mindsets, attribution for failure, expectancy for success, interest in math and anxiety about math, and belonging uncertainty

Data for these variables are from student record data and student responses to survey questions collected in the first online session before they started the training materials.

Second, a composite measure of schools' overall academic achievement level before the study was used to stratify the national population of high schools for sample-selection purposes and was also used for school subgroup definitions, discussed in the report. This composite measure was constructed by the NSLM team using information from nationally available school-performance databases.

Third, a set of school characteristics such as enrollment, racial and socioeconomic composition of students, school location, and student-teacher ratio was used to assess whether the findings based on the analytic sample can be generalized to the targeted national population of high schools. Such information was collected from the Common Core of Data, a publicly available database that contains school-level information for all schools in the nation.

Fourth, a set of variables constructed from students' responses to questions and tasks immediately after each of the online sessions was used to measure their mindset beliefs and challenge-seeking intentions and behaviors. These were considered intermediate outcomes of the program.

Last, students' GPAs for core ninth-grade courses in math, English/language arts, science, and social science were used to construct the pre- and postprogram academic measures for the study. Specifically, the team constructed the following three academic outcomes using data described in the last section:

1. *Postprogram average GPA*: This is the average GPA across the four core ninth-grade subject areas and measures students' general academic performance level.
2. *Poor-performance indicator*: This is a dichotomous indicator that equals 1 if the postprogram average grade is 1.0 or lower and equals zero otherwise.
3. *Postprogram math GPA*: Math GPA is singled out because many researchers consider math to be a gateway course for high school success.

To assess the program's potential impacts on students' performance in standardized, high-stakes tests; their academic engagement; and their behavior, the NSLM research team also attempted to collect students' ninth-grade state test data, their attendance data, and their office disciplinary record data. However, because the response rates for these data are very low (generally below 50 percent), the MDRC evaluation did not include them in the analysis for this report.

Appendix B

Estimation Methods and Sample Characteristics

The National Study of Learning Mindsets (NSLM) study uses an experimental design that randomly assigns individual ninth-grade students in a school to a program group, who received the growth mindset intervention, and to a control group, who did not receive the intervention. The control group serves as a benchmark, or “counterfactual,” for how students in the program group would have performed if they had not experienced the intervention. Therefore, the impacts (the differences in outcomes between the program and control groups) represent the effects that the intervention had on student outcomes over and above what the students would have achieved had they not been exposed to the intervention.

As stated in the preregistered analysis plan for the NSLM, the impact of the growth mindset intervention is estimated for each outcome using the following statistical model:

$$Y_{ij} = \sum_{j=1}^{j} \alpha_j \cdot S_{ij} + \beta \cdot T_{ij} + \sum_{k=1}^{K} \theta_k \cdot X_{kij} + e_{ij} \qquad \text{(B.1)}$$

Where:

Y_{ij} = the outcome for student *i* in school *j*,

S_{ij} = indicator variable indicating student *i* attended school *j*,

T_{ij} = 1 if student *i* in school *j* was randomized to the program group and zero otherwise,

X_{kij} = baseline covariate *k* for student *i* in school *j*,

e_{ij} = student-level random error term for student *i* in school *j*, assumed to be normally distributed with mean zero and variance of σ^2.

The model is estimated with student-level survey weights that account for school-level and student-level adjustments for sampling probability and nonresponse. It uses cluster-robust standard errors, clustered at the school level, so that standard errors appropriately account for the uncertainty associated with generalizing from the sample of schools in the study to the population of the school from which they were randomly selected.

The coefficient β therefore represents the overall average impact of being randomized to the program instead of the control condition *for all ninth-grade students in the national population of high schools targeted by this study*. The t-statistic for this coefficient tests whether the estimated average impact for students in the national high school population identified for this study is different from zero to a statistically significant degree. Similar models are used in the analyses for all students as well as for the subgroups.

There are several features to note about this model:

- Weighted ordinary least squares (OLS) regression is used to estimate Equation B.1.
- Indicators for random assignment blocks (school in this case) are included in the model to reflect the design feature (that is, differential rates of research

group assignment by block) and to control for variation in mean outcome levels across schools (which can be due to different characteristics of schools or their students, for example).

- The model controls for the students' preprogram achievement scores, including average grade point average (GPA) for the four core subjects and math GPA. Doing so controls for baseline differences that might have occurred by chance and increases the precision of impact estimates because pretests substantially reduce within-school random error in the outcome measure.

- The model also controls for students' preprogram mindset measures, including students' overall mindset beliefs, attributions for failure, expectancy for success, interest in math, math anxiety, and belonging uncertainty. These covariates measure student mindsets before the program, and to the extent that one's mindset is associated with academic performance, controlling for these variables can potentially reduce the random error in the outcomes.

- Other baseline covariates are added to the model to improve precision. These covariates include students' gender, race/ethnicity, free or reduced-price lunch (FRPL) status, and special education status.

- To keep the analysis sample as complete as possible, the missing values for a given covariate are imputed as zero, and a dummy variable indicating whether a student is missing this covariate or not is also included in the regression.

School Sample and Generalizability

The primary goal of the NSLM study is to test the effectiveness of the growth mindset intervention in the population of ninth-grade students attending regular public high schools with at least 25 ninth-graders and with ninth grade as the lowest grade in the school. To achieve this goal, the NSLM research team identified 11,221 eligible high schools based on their organizational type, enrollment size, and grade configuration.[1] The NSLM research team then randomly selected a sample of 139 high schools for recruitment into the study. Of these 139 high schools, 63 (45 percent) agreed to participate in the NSLM study and also provided the research team with usable school record data. Ninth-grade students from these 63 schools constitute the analytic sample for the MDRC evaluation.

Since not all selected schools agreed to participate in or to provide data to the study, it is important to assess whether the findings based on this analytic sample are generalizable to the targeted national population of eligible high schools (hereafter referred to as the inference population). Two conditions have to hold for the findings *not* to be generalizable to the inference population: First, the study sample must be systematically different from the target

[1]See Gopalan and Tipton (2018) for eligibility criteria.

population, in either observable characteristics for which the team has data or unobservable characteristics that cannot be captured with available data. Second, the program impacts must vary across schools so that the average program effects for the schools in the study sample differ from those of the schools that are not in the study sample. Because outcome measures for all schools in the inference population are not available, it is not feasible to test the second condition. However, it is possible to assess the first condition by comparing observable school characteristics between the analytic sample and the inference population (that is, the 11,221 eligible regular high schools). These characteristics were collected by the NSLM research team through publicly available databases such as the Common Core of Data.

Appendix Table B.1 presents the results of these comparisons and shows that the differences between these two samples are small across most variables available for examination. The 63 schools in the analytic sample are very similar to the inference population in terms of the two key criteria used in the sample selection process: school achievement level and proportion of minority students in the school. Not only are these estimated differences not statistically significant at the 5 percent level, but their absolute magnitudes range from 0.04 to 0.18 standard deviations in effect size, smaller than the 0.25 standard deviations threshold set forward by the latest What Works Clearinghouse standard for substantial baseline differences.[2] The analytic sample is also similar to the inference population in terms of the proportion of students with poverty status, the student-to-teacher ratio, and overall and ninth-grade enrollments. The only place where the two groups differ significantly is in terms of school location: Compared with schools in the analytic sample, schools in the national population are more likely to be in suburban areas instead of urban areas.

These results show that despite some schools' nonparticipation and nonresponse, for a set of observed school characteristics, the analytic sample is mostly similar to the inference population. However, it is still possible that they differ in unobservable characteristics such as teacher or leadership quality, and these characteristics could affect how the intervention works in the schools. Therefore, in order to generalize the findings based on the analytic sample to the inference population, one has to assume either that the sample schools are similar to the national population in unobserved characteristics or that the program impacts do not vary by the unobserved characteristics that are different between the sample schools and the inference population. If either of these two assumptions holds, then the impact findings based on the analytic sample can be generalized to the national population of regular public high schools with at least 25 ninth-graders and serving ninth through twelfth grade.

Student Sample and Internal Validity

The student sample of MDRC's evaluation consists of 11,888 ninth-grade students from 63 high schools around the country. Among these students, 5,916 (49.8 percent) were randomly assigned to the program group, while 5,972 (50.2 percent) were randomly assigned to the

[2] What Works Clearinghouse (2017).

control group. This sample is limited to students with valid GPA scores at the end of ninth grade. This random assignment design creates the expectation that students in the program and control groups were similar on average before the intervention and that the differences in their outcomes can be attributed to the effects of the enhanced program.

Appendix Table B.2 shows that the program and control group students are similar to each other for many of the demographic characteristics. Specifically, for both groups, about 45 percent to 46 percent of the students are white, non-Hispanic; about 50 percent are male; 13 percent have special education status; and about 55 percent to 56 percent are eligible for free or reduced-price lunch. In addition, students in these two groups are similar in their responses to survey questions related to their mindset beliefs and other psychological attributes before the intervention, as shown in the bottom panel of the table.

However, differences exist in the preprogram academic performances between these two groups. Specifically, the program group had a lower average GPA than the control group before the start of the program: The estimated difference in the preprogram average GPA between these two groups is 0.05 points (effect size = 0.05, p-value = 0.037); similarly, the estimated difference in preprogram math GPA is 0.06 points (effect size = 0.05, p-value = 0.034). Measures of student characteristics (including students' preprogram average and math GPAs) are included in the impact model in order to control for these observed differences.

An omnibus test was conducted to see if there is a systematic difference between the program and control groups across all background characteristics listed in Appendix Table B.2. Results from that test indicate no such difference in the background characteristics of students in the program and the control groups (p-value = 0.654), and the statistical equivalence of the two research groups is largely preserved in the sample used for the analysis.

Note that the results reported in Appendix Table B.2 are based on weighted sample means from both groups, where the weights account for sampling and nonresponse at both the school level and the student level. The results without the weights are similar to those in the table, except that the difference between the program and control groups in the preprogram average GPA is -0.03 points (effect size = -0.03, p-value = 0.11), and the difference is -0.04 points (effect size = -0.04, p-value = 0.05) for preprogram math GPA. The F-test for systematic difference between the two samples yields a p-value of 0.99.

Nonetheless, a series of sensitivity checks were conducted to examine whether the difference in students' achievement level before the program might affect the impact estimation. Specifically, four to eight schools with the largest differences in prior average GPA between students in the program and control groups were sequentially dropped from the analysis so that in the remaining sample of students, there was no longer a significant difference between students in the program and control groups in their preprogram average GPA. All impacts were then reestimated using these restricted samples where two groups were similar in their prior academic performance levels. In general, impact estimates based on the restricted samples are

similar in magnitude and statistical significance levels to those estimates for the full sample presented in the report (Appendix Table B.3). These results demonstrate that the impact estimates are not affected by the difference in preprogram achievement levels between the program and control group students.

Appendix Table B.1

Characteristics of Schools in the Analytic Sample and the Inference Population

Characteristics	Analytic Sample Weighted Mean	Population Mean	Difference	Difference in Effect Size	P-Value of Difference
School previous achievement level	-0.04	0.00	-0.04	-0.04	0.777
Proportion of minority students (%)	32.31	38.23	-5.92	-0.18	0.160
Proportion of students with poverty status (%)	19.32	20.31	-0.99	-0.10	0.450
Student-teacher ratio	16.51	16.16	0.36	0.07	0.596
Total enrollment	996	1,032	-36.04	-0.05	0.699
Ninth-grade enrollment	263	284	-20.61	-0.10	0.445
School locale (%)					
Urban	36.66	22.84	13.82 **	0.33	0.017
Suburban	11.85	27.32	-15.47 **	-0.35	0.012
Town	24.54	17.62	6.92	0.18	0.180
Rural	26.94	32.22	-5.28	-0.11	0.402
Number of schools	63	11,221			

SOURCE: School information collected by ICF International before random assignment.

NOTES: The first column reports the mean values of school characteristics for the analytic sample schools, weighted at the school-level account for sampling probability and data availability. The second column reports the mean values of school characteristics for the inference population of all eligible regular high schools in the nation. The statistical significance of each difference between the analytic sample and the inference population was tested with a two-tailed t-test. Statistical significance level is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Appendix Table B.2

Background Characteristics of All Students in the Program and Control Groups

Characteristics	Program Group	Control Group	Estimated Difference	Estimated Difference in Effect Size	Standard Error for Estimated Difference	P-Value for Estimated Difference
Male (%)	50.20	50.23	-0.03	0.00	0.01	0.976
Race/ethnicity (%)						
Hispanic	20.97	20.12	0.86	0.02	0.01	0.227
Black, non-Hispanic	12.38	11.85	0.52	0.02	0.01	0.335
White, non-Hispanic	45.05	45.89	-0.84	-0.02	0.01	0.438
Other	21.60	22.14	-0.54	-0.01	0.01	0.524
With special education status (%)	13.24	13.19	0.05	0.00	0.01	0.951
With poverty status (%)	55.49	56.10	-0.61	-0.01	0.01	0.455
Student preprogram academic performance						
Average GPA	2.80	2.84	-0.05 **	-0.05	0.02	0.037
Math GPA	2.70	2.76	-0.06 **	-0.05	0.03	0.034
Student initial survey responses						
Overall mindset belief	2.73	2.74	-0.02	-0.01	0.03	0.542
Attributions for failure	2.08	2.10	-0.02	-0.01	0.02	0.458
Expectancy for success	5.15	5.18	-0.04	-0.03	0.03	0.171
Interest in math	2.66	2.67	-0.01	-0.01	0.02	0.600
Math anxiety	2.54	2.54	0.01	0.00	0.03	0.836
Belonging uncertainty	2.12	2.09	0.02	0.02	0.02	0.343

SOURCES: Student survey responses and student record data from school years 2014-2015 and 2015-2016.

NOTES: Students' poverty status is determined by their eligibility for the free or reduced-price lunch (FRPL) program. GPA is grade point average, measured on a scale of 0.0 to 4.3. The sample includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The number of observations ranges from 6,372 to 11,868 due to varying rates of missingness for each variable.

The estimated differences are regression-adjusted using ordinary least squares (OLS) regressions that account for the random assignment blocks (schools). The estimated standard errors are adjusted to account for the clustering of students within schools. Student-level weights that adjust for sampling probability and data availability at both the student and school levels are applied to all regressions. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

A chi-square test for joint significance across all variables yields a p-value of 0.654.

Appendix Table B.3

Robustness Checks for Impact Findings for All Students

	Benchmark		Trimming Exercise: Dropping Schools with the Largest Preprogram GPA Differences					
			Dropping Four Schools		Dropping Six Schools		Dropping Eight Schools	
Outcomes	Estimated Impact	P-Value for Estimated Impact	Estimated Impact	P-Value for Estimated Impact	Estimated Impact	P-Value for Estimated Impact	Estimated Impact	P-Value for Estimated Impact
Postprogram average GPA (0-4.3 scale)	0.05 ***	0.004	0.04 ***	0.006	0.04 ***	0.002	0.05 ***	0.001
Poor performance (%, GPA of 1.0 or lower)	-2.37 ***	0.005	-1.89 ***	0.010	-2.17 ***	0.002	-2.25 ***	0.002
Postprogram math GPA (0-4.3 scale)	0.06 ***	0.004	0.06 ***	0.007	0.06 ***	0.004	0.07 ***	0.002
Number of students	11,888		11,511		11,309		11,050	
Number of schools	63		59		57		55	

SOURCES: Student survey responses collected during the online sessions and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average. The sample for the top panel includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The sample size is 11,888 for average GPA and poor performance indicator and 10,853 for math GPA. Sample size varies for the trimming exercises.

The estimated impacts are regression-adjusted using ordinary least squares (OLS) regressions that account for the random assignment blocks (schools). The model used for the benchmark and the trimming excercises also controls for students' preprogram characteristics including their demographics, initial mindset beliefs and attitudes, and preprogram academic achievement levels. The estimated standard errors are adjusted to account for the clustering of students within schools. Student-level weights that adjust for sampling probability and data availability at both the student and school levels are applied to all regressions. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Appendix C

Supplementary Tables for Impact Findings

Impact Findings for All Students

Appendix Table C.1 provides details for the impact estimates on students' mindset beliefs, challenge-seeking behaviors, and academic outcomes for all students.

Impact Findings for Student Subgroups

Not all students in the analytic sample have their preprogram average grade point averages (GPAs) reported by their schools. In fact, about 9 percent of all students have missing values for this variable. They were excluded from the estimations of the impacts for lower- and higher-performing student subgroups reported in Figure 4 in the report. It is possible that those students with missing preprogram GPAs were different from those with nonmissing data and the subgroup impact findings reported in Figure 4 might be misleading.

To address this concern, the MDRC team used other available preprogram characteristics of the students in a profile analysis to identify whether the students with missing preprogram GPAs should be placed in the lower- or higher-performing student subgroups. All students with preprogram GPAs were assigned to the subgroups based on their GPA values as before, and only those with missing preprogram GPAs were assigned to the subgroups based on results from the profile analysis. The top panel in Appendix Table C.2 presents the impact estimates for the lower- and higher-performing student subgroups based on this alternative subgroup definition based on partially imputed data. It shows that the subgroup impact patterns reported in Figure 4 do not change substantively.

The rest of Appendix Table C.2 presents impact findings for other student subgroups discussed in the report, including subgroups based on students' initial fixed mindset rating, race/ethnicity, gender, or poverty status. The results show that the program impacts did not vary across these subgroups.

Impact Findings for School Subgroups

The preregistered analysis plan proposed to use a random effect model to assess the variability in the impacts across schools.[1] This two-level model is specified as the following:

Level 1 (student level)

$$Y_{ij} = \sum_{j=1}^{j} \alpha_j \cdot S_{ij} + \beta_j \cdot T_{ij} + \sum_{k=1}^{K} \theta_k \cdot X_{kij} + e_{ij} \tag{C.1}$$

Level 2 (school level)

$$\beta_j = \beta + \gamma_j \tag{C.2}$$

[1]The model is based on Bloom, Raudenbush, Weiss, and Porter (2017).

Where all variables are defined as in Equation B.1 and

γ_j = school-level random error term for school j, assumed to be normally distributed with mean zero and variance of τ^2.

Usually the impact variation measure, τ^2, can be estimated using PROC MIXED in SAS or MIXED in STATA. The estimated τ^2 captures the variance of the school-level impact estimates. A chi-square test based on Q-statistics calculated using school-level impact estimates can be used to ascertain the statistical significance of τ^2. This approach is widely used in meta-analysis to test the null hypothesis of zero cross-study impact variation.[2]

However, due to convergence issues potentially caused by the specific data structure of the sample, the two-level model described in Equations C.1 and C.2 cannot be estimated. Therefore, the magnitude of the impact variation as measured by τ^2 cannot be estimated. However, the team was still able to calculate the Q-statistics based on school-level impact estimates. Appendix Table C.3 shows the Q-statistics and corresponding p-values from the chi-square tests for all students and for the lower-performing students for the three academic outcomes. These test results show that by and large, the variation in school-level impacts is not statistically significant. This is true for both samples of students and for all three academic outcomes.

The statistical significance of cross-school impact variation should not be used as a "gateway" test of whether to attempt to predict variation in the program effects, because an omnibus test of whether estimated effects vary across sites (like the chi-square test) can have less power (sometimes far less power) than a focused test of the relationship between the effects and a specific school-level characteristic or moderator.[3]

This study explored such relationships through school-level subgroup analysis and focused on two types of school characteristics that were specified in the preregistered analysis plan: school achievement levels and school growth mindset climate. The latter factor was measured in two different ways: by school average fixed mindset ratings and by the prevalence of challenge-seeking behaviors in school. The analyses estimated separate impacts for each of these subgroups and tested for the statistical significance of the difference in subgroup impacts. Appendix Table C.4 presents results based on these analyses.

[2]Hedges and Olkin (2014).

[3]See Bloom, Raudenbush, Weiss, and Porter (2017), Appendix C.

Appendix Table C.1

Estimated Impacts on Student Mindsets, Challenge-Seeking Behaviors, and Academic Outcomes for All Students

Measures	Program Group	Control Group	Estimated Impact	Estimated Impact in Effect Size	Standard Error for Estimated Impact	P-Value for Estimated Impact
Student mindset attitudes and beliefs						
Overall fixed mindset level (1-6 scale)	2.26	2.60	-0.34 ***	-0.26	0.02	0.000
You have a certain amount of intelligence, and you really can't do much to change it.	2.23	2.59	-0.36 ***	-0.26	0.02	0.000
Your intelligence is something about you that you can't change very much.	2.29	2.61	-0.31 ***	-0.22	0.03	0.000
Being a math person or not is something you can't really change; some people are good at math and other people aren't.	2.98	3.51	-0.52 ***	-0.36	0.03	0.000
One of my main goals for the rest of the school year is to avoid looking dumb in my classes.	3.71	3.86	-0.14 ***	-0.10	0.04	0.001
Attributions for failure (1-6 scale)	3.61	3.66	-0.05 **	-0.06	0.02	0.026
This means I am probably not very good at math.	2.67	2.78	-0.11 ***	-0.08	0.03	0.000
I can get a higher score next time if I find a better way to study.[a]	4.54	4.53	0.01	0.01	0.03	0.831
Challenge-seeking intentions (%)	50.80	38.19	12.60 ***	0.26	1.07	0.000
Challenge-seeking behaviors						
Number of hard items chosen	3.33	2.83	0.50 ***	0.21	0.06	0.000
Number of easy items chosen	3.54	4.00	-0.46 ***	-0.18	0.06	0.000
Categorical measures (%)						
More hard items than easy ones	39.44	30.64	8.81 ***	0.19	1.12	0.000
Equal number of hard and easy items	12.83	12.54	0.30	0.01	0.57	0.604
More easy items than hard ones	47.72	56.82	-9.10 ***	-0.18	1.14	0.000
Mean item challenge (1-3 scale)	1.96	1.86	0.10 ***	0.23	0.01	0.000
Total challenge value across items	21.03	19.94	1.09 ***	0.11	0.25	0.000

(continued)

Appendix Table C.1 (continued)

Measures	Program Group	Control Group	Estimated Impact	Estimated Impact in Effect Size	Standard Error for Estimated Impact	P-Value for Estimated Impact
Academic outcomes						
Postprogram GPA	2.59	2.55	0.05 ***	0.04	0.02	0.004
Poor performance (%, GPA of 1.0 or lower)	29.72	32.09	-2.37 ***	-0.05	0.01	0.005
Postprogram math GPA	2.48	2.42	0.06 ***	0.05	0.02	0.004

SOURCES: Student survey responses collected during the online sessions and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average, measured on a scale of 0.0 to 4.3. The sample includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The sample size ranges from 10,007 to 11,888 due to varying rates of missing values for each outcome measure.

[a]This variable is reverse coded so that higher values indicate stronger fixed mindset beliefs.

The estimated impacts are regression-adjusted using ordinary least squares (OLS) regressions that account for the random assignment blocks (schools). The model also controls for students' baseline characteristics including their demographics, initial mindset, and preprogram academic achievement levels. The estimated standard errors are adjusted to account for the clustering of students within schools. Student-level weights that adjust for sampling probability and data availability at both the student and school levels are applied to all regressions. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Appendix Table C.2

Impacts on Academic Performance for Student Subgroups

Outcome	Program Group	Control Group	Estimated Impact	Estimated Impact in Effect Size	Standard Error for Estimated Impact	P-Value for Estimated Impact	Estimated Subgroup Difference in Impact	P-Value for Estimated Subgroup Difference in Impact
Alternative subgroups by preprogram average GPA								
Postprogram GPA							-0.06 †††	0.006
Lower	2.11	2.03	0.07 ***	0.07	0.02	0.000		
Higher	3.16	3.15	0.01	0.01	0.02	0.596		
Poor performance (%, GPA of 1.0 or lower)							3.99 †††	0.001
Lower	45.36	49.29	-3.93 ***	-0.09	1.18	0.001		
Higher	11.65	11.58	0.06	0.00	0.71	0.927		
Postprogram math GPA							-0.04	0.461
Lower	1.99	1.91	0.08 ***	0.06	0.03	0.006		
Higher	3.06	3.01	0.04	0.03	0.04	0.243		
Subgroups by preprogram fixed mindset rating								
Postprogram GPA							-0.02	0.581
Lower	2.71	2.68	0.04 **	0.03	0.02	0.035		
Higher	2.44	2.38	0.06 *	0.05	0.03	0.077		
Poor performance (%, GPA of 1.0 or lower)							0.32	0.830
Lower	25.20	27.29	-2.09 **	-0.05	0.93	0.025		
Higher	35.68	38.09	-2.41 *	-0.05	1.28	0.060		
Postprogram math GPA							-0.05	0.324
Lower	2.57	2.53	0.04	0.03	0.03	0.134		
Higher	2.36	2.28	0.09 **	0.07	0.04	0.027		

(continued)

Appendix Table C.2 (continued)

Outcome	Program Group	Control Group	Estimated Impact	Estimated Impact in Effect Size	Standard Error for Estimated Impact	P-Value for Estimated Impact	Estimated Subgroup Difference in Impact	P-Value for Estimated Subgroup Difference in Impact
Subgroups by race/ethnicity								
Postprogram GPA							0.00	0.920
White	2.89	2.84	0.05 *	0.05	0.03	0.088		
Other	2.36	2.31	0.05 *	0.04	0.02	0.056		
Poor performance (%, GPA of 1.0 or lower)							-1.31	0.581
White	20.45	22.06	-1.61	-0.03	1.71	0.347		
Other	37.19	40.11	-2.92 **	-0.06	1.16	0.012		
Postprogram math GPA							0.00	0.972
White	2.74	2.68	0.06 **	0.05	0.03	0.014		
Other	2.25	2.19	0.07 *	0.05	0.04	0.084		
Subgroups by gender								
Postprogram GPA							0.00	0.910
Female	2.84	2.79	0.05 *	0.04	0.03	0.078		
Male	2.36	2.31	0.05 ***	0.05	0.01	0.000		
Poor performance (%, GPA of 1.0 or lower)							-0.10	0.937
Female	21.62	24.06	-2.44 **	-0.05	1.21	0.044		
Male	37.64	40.18	-2.54 ***	-0.05	0.84	0.002		
Postprogram math GPA							0.01	0.790
Female	2.72	2.66	0.06 *	0.05	0.03	0.072		
Male	2.25	2.18	0.07 ***	0.05	0.02	0.002		

(continued)

Appendix Table C.2 (continued)

Outcome	Program Group	Control Group	Estimated Impact	Estimated Impact in Effect Size	Standard Error for Estimated Impact	P-Value for Estimated Impact	Estimated Subgroup Difference in Impact	P-Value for Estimated Subgroup Difference in Impact
Subgroups by poverty status								
Postprogram GPA							-0.06	0.201
With poverty status	2.20	2.20	0.00	0.00	0.02	0.835		
Without poverty status	2.90	2.83	0.06 *	0.06	0.04	0.091		
Poor performance (%, GPA of 1.0 or lower)							0.78	0.629
With poverty status	42.51	43.87	-1.36	-0.03	1.06	0.198		
Without poverty status	20.22	22.36	-2.14	-0.05	1.46	0.144		
Postprogram math GPA							-0.13	0.135
With poverty status	2.12	2.13	-0.02	-0.01	0.05	0.702		
Without poverty status	2.75	2.64	0.11 **	0.09	0.06	0.040		

SOURCES: Student survey responses collected during the online sessions and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average,measured on a scale of 0.0 to 4.3. A student's poverty status is determined by his or her eligibility for the free or reduced-price lunch (FRPL) program. The sample includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The sample size varies due to varying rates of missing values for each variable used to define the subgroups: Students' preprogram average GPA, preprogram fixed mindset rating, race, gender, and poverty status were used to define the subgroups.

The estimated impacts are regression-adjusted using ordinary least squares (OLS) regressions that account for the random assignment blocks (schools). The model also controls for students' baseline characteristics including their demographics, parental education, initial mindset, and baseline academic achievement levels. The estimated standard errors are adjusted to account for the clustering of students within schools. Student-level weights that adjust for sampling probability and data availability at both the student and school levels are applied to all regressions. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Two-tailed t-tests were used to test differences in estimated impacts between the two subgroups. Statistical significance is indicated by the following: ††† denotes a p-value < 0.01, †† denotes a p-value < 0.05, and † denotes a p-value < 0.10.

Appendix Table C.3

Significance Tests for Cross-School Impact Variation

	All Students		Lower-Performing Students	
Outcome	Q-Statistic	P-Value	Q-Statistic	P-Value
Postprogram GPA	65.84	0.346	47.07	0.905
Poor performance (%, GPA of 1.0 or lower)	74.65	0.130	47.89	0.889
Postprogram math GPA	58.19	0.614	50.52	0.804

SOURCES: Student survey responses collected during the online sessions and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average, measured at a scale of 0.0 to 4.3. The all-student sample includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The sample size ranges from 10,853 for math GPA to 11,888 for the other two outcomes due to missing values.

Students whose preprogram average GPA was below school-level median GPA are in the lower-performing subgroup. The lower-performing student sample includes 5,503 students for the average GPA and poor-performance indicator analysis and 4,992 students for the math GPA analysis.

Q-statistics were calculated based on school-level impact estimates. P-values are based on chi-square tests. Statistical significance is indicated as the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Appendix Table C.4

School-Level Subgroup Impact Estimates for All Students and for Lower-Performing Students, by School Subgroups

Outcome	All Students: Estimated Impact	All Students: Standard Error of Estimated Impact	All Students: P-Value of Estimated Impact		Lower-Performing Students: Estimated Impact	Lower-Performing Students: Standard Error of Estimated Impact	Lower-Performing Students: P-Value of Estimated Impact	
Subgroup by school academic achievement level								
Postprogram GPA								
Low	0.00	0.04	0.937		0.04	0.04	0.398	
Medium	0.06	0.02	0.006	***	0.08	0.03	0.027	**
High	0.01	0.02	0.433		0.02	0.02	0.165	
Difference between medium and other groups	0.05	0.03	0.111		0.04	0.04	0.278	
Poor performance (%, GPA of 1.0 or lower)								
Low	-1.89	2.37	0.429		-1.81	2.52	0.476	
Medium	-3.45	1.10	0.003	***	-5.02	1.90	0.010	**
High	0.50	1.14	0.664		0.25	0.91	0.788	
Difference between medium and other groups	-2.84	1.57	0.076	†	-4.28	2.20	0.056	†
Postprogram math GPA								
Low	0.01	0.07	0.862		0.12	0.06	0.065	*
Medium	0.09	0.02	0.000	***	0.09	0.04	0.013	**
High	0.00	0.04	0.915		-0.01	0.05	0.914	
Difference between medium and other groups	0.09	0.04	0.027	††	0.05	0.05	0.308	
Subgroup by school average fixed mindset belief								
Postprogram GPA								
Weak fixed mindset belief	0.04	0.02	0.060	*	0.06	0.04	0.083	*
Strong fixed mindset belief	0.04	0.02	0.022	**	0.05	0.02	0.025	**
Subgroup difference	0.00	0.03	0.956		-0.02	0.04	0.702	
Poor performance (%, GPA of 1.0 or lower)								
Weak fixed mindset belief	-2.63	1.27	0.043	**	-3.65	1.89	0.058	*
Strong fixed mindset belief	-1.96	0.94	0.041	**	-3.12	1.65	0.063	*
Subgroup difference	0.67	1.58	0.671		0.53	2.51	0.834	
Postprogram math GPA								
Weak fixed mindset belief	0.05	0.03	0.065	*	0.05	0.04	0.219	
Strong fixed mindset belief	0.07	0.03	0.027	**	0.08	0.03	0.018	**
Subgroup difference	0.01	0.04	0.752		0.03	0.05	0.634	

(continued)

Appendix Table C.4 (continued)

	All Students			Lower-Performing Students		
Outcome	Estimated Impact	Standard Error of Estimated Impact	P-Value of Estimated Impact	Estimated Impact	Standard Error of Estimated Impact	P-Value of Estimated Impact
Subgroup by prevalence of challenge-seeking behavior in school						
Postprogram GPA						
Low prevalence	-0.01	0.02	0.395	0.00	0.02	0.940
High prevalence	0.09	0.02	0.000 ***	0.11	0.03	0.001 ***
Subgroup difference	0.10	0.03	0.000 †††	0.11	0.04	0.006 †††
Poor performance (%, GPA of 1.0 or lower)						
Low prevalence	0.33	0.90	0.715	-0.40	1.08	0.713
High prevalence	-4.39	1.10	0.000 ***	-5.82	1.87	0.003 ***
Subgroup difference	-4.72	1.42	0.002 †††	-5.42	2.16	0.015 ††
Postprogram math GPA						
Low prevalence	-0.02	0.02	0.426	-0.01	0.04	0.868
High prevalence	0.12	0.03	0.000 ***	0.12	0.03	0.000 ***
Subgroup difference	0.13	0.03	0.000 †††	0.13	0.05	0.017 ††

SOURCES: Student survey responses collected during the online sessions and student record data from school years 2014-2015 and 2015-2016.

NOTES: GPA is grade point average, measured on a scale of 0.0 to 4.3. The all-student sample includes ninth-grade students in the 63 schools for whom ninth-grade achievement information is available. The sample size is 11,888 for the average GPA and poor-performance indicator analysis and 10,853 for math GPA analysis. Students whose preprogram GPAs were below the school-level median GPA are in the lower-performing student subgroup. The lower-performing student sample includes 5,503 students for the average GPA and poor-performance indicator analysis and 4,992 students for the math GPA analysis.

Subgroups are defined based on school academic achievement level, school average preprogram fixed mindset rating, and prevalence of challenge-seeking behaviors in school as measured by the average hard items chosen in the “Make a Math Worksheet” task by control group students in the school.

The estimated impacts are regression-adjusted using ordinary least squares (OLS) regressions that account for the random assignment blocks (schools). The model also controls for students’ baseline characteristics including their demographics, parental education, initial mindset, and baseline academic achievement levels. The estimated standard errors are adjusted to account for the clustering of students within schools. Student-level weights that adjust for sampling probability and data availability at both the student and school levels are applied to all regressions. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by the following: *** denotes a p-value < 0.01, ** denotes a p-value < 0.05, and * denotes a p-value < 0.10.

Two-tailed t-tests were used to test differences in estimated impacts between the two subgroups. Statistical significance is indicated by the following: ††† denotes a p-value < 0.01, †† denotes a p-value < 0.05, and † denotes a p-value < 0.10.

Appendix D

Comparing Key Impact Findings with the National Study of Learning Mindsets (NSLM) Results

The key impact findings presented in this report are substantively consistent with the findings reported by the NSLM research team.[1] Appendix Table D.1 compares the impact findings for a set of common outcomes and similar samples reported in these two studies.

Several factors might have contributed to the observed small differences in the estimated impacts between the two studies. First, the samples of schools used by the two studies differ slightly: The MDRC study excluded two schools with vague grade information from the analysis, while the NSLM team strictly followed the preregistered analysis plan and included these schools in its analysis. Second, the set of covariates in the impact model was not exactly the same in these two studies: There are slight differences in the way categorical variables were coded, and in the inclusion or exclusion of certain variables. Third, the weights used in the estimation differ slightly between the two studies: The MDRC team used trimmed weights that put a cap on extremely large probability weights, while the NSLM team used the prespecified probability weights. Last, as mentioned in the report, the MDRC team did not include students with missing preprogram grade point average (GPA) information (about 9 percent of the total sample) in the primary student achievement subgroup analysis, while the NSLM team used imputed preprogram information to define subgroup status for these students with missing preprogram information. Consequently, the student sample for the student achievement subgroup analysis differs slightly between these two studies.

Despite these differences in samples and estimation model specification, the results presented in this report and summarized in Appendix Table D.1 demonstrate that the findings reported by the NSLM team are robust to these reasonable deviations from the preregistered analysis.

[1]Yeager et al. (2019).

Appendix Table D.1

Comparison of Findings Between the Current Study and Yeager et al. (2019)

Finding	Yeager et al. (2019)	MDRC Study
On all students' mindset beliefs	The intervention reduced fixed mindset beliefs as measured by mindset rating (impact = -0.42).	The intervention reduced fixed mindset beliefs as measured by responses to a set of survey questions (impacts range from -0.52 to -0.31).
On all students' academic achievement	Effect on average grade point average (GPA) is 0.05 points; effect on math GPA is 0.06 points.	Effect on average GPA is 0.05 points; effect on math GPA is 0.06 points.
On lower-performing students' academic achievement	Effect on average GPA is 0.10 points; effect on math GPA is 0.09 points.	Effect on average GPA is 0.06 points, effect on math GPA is 0.06 points.
On lower-performing students in schools with different achievement levels	• Effects on ninth-grade GPAs are statistically smaller for schools with higher achievement levels. • Effects for medium-achieving schools are similar to those for low-achieving schools, but larger than those for high-achieving schools.	• Effects for medium-achieving schools are beneficial and statistically significant for all three academic outcomes. • Effects for medium-achieving schools are different from other schools for one of the three academic outcomes.
On lower-performing students in schools with low or high prevalence of challenge-seeking behaviors	Effects are larger for schools with high prevalence.	• Effects for schools with high prevalence are positive and statistically significant. • Effects are different between schools with low and high prevalence.
On lower-performing students in schools with weak or strong initial beliefs of fixed mindset	Effects are not different between the two sets of schools.	Effects are not different between the two sets of schools.

References

Allensworth, Elaine, and John Easton. 2005. *The On-Track Indicator as a Predictor of High School Graduation.* Chicago: University of Chicago Consortium on Chicago School Research.

Aronson, Joshua, Carrie Fried, and Catherine Good. 2002. "Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence." *Journal of Experimental Social Psychology* 38, 2: 113-125.

Blackwell, Lisa, Kali Trzesniewski, and Carol Dweck. 2007. "Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention." *Child Development* 78, 1: 246-263.

Bloom, Howard. S., Stephan W. Raudenbush, Michael J. Weiss, and Kristin Porter. 2017. "Using Multisite Experiments to Study Cross-Site Variation in Effects of Program Assignment." *Journal of Research on Educational Effectiveness* 10, 4: 817-842.

Boulay, Beth, Barbara Goodson, Rob Olsen, Rachel McCormick, Catherine Darrow, Michael Frye, Katherine Gan, Eleanor Harvill, and Maureen Sarna. 2018. *The Investing in Innovation Fund: Summary of 67 Evaluations: Final Report*. NCEE 2018-4013. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bradby, Denise, Rosio Pedroso, and Andy Rogers. 2007. *Secondary School Course Classification System: School Codes for the Exchange of Data (SCED)*. NCES 2007-341. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Duckworth, Angela L., and David Yeager. 2015. "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes." *Educational Researcher* 44, 4: 237-251.

Dweck, Carol S., and Ellen L. Leggett. 1988. "A Social-Cognitive Approach to Motivation and Personality." *Psychological Review* 95, 2: 256.

Goldin, Claudia, and Lawrence F. Katz. 2008. "Transitions: Career and Family Life Cycles and the Educational Elite." *American Economic Review* 98, 2: 363-369.

Good, Catherine, Joshua Aronson, and Michael Inzlicht. 2003. "Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat." *Journal of Applied Developmental Psychology* 24, 6: 645-662.

Gopalan, M., and Elizabeth Tipton. 2018. "Is the National Study of Learning Mindsets Nationally Representative?" Working paper. PsyArXiv, Nov. 3. Website: https://psyarxiv.com/dvmr7.

Hanushek, Eric A., and Ludger Woessmann. 2008. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature* 46, 3: 607-668.

Hedges, Larry V., and Olkin, I. 2014. *Statistical Methods for Meta Analysis*. San Diego: Academic Press.

Hochberg, Yosef, and Yoav Benjamini. 1990. "More Powerful Procedures for Multiple Significance Testing." *Statistics in Medicine* 9, 7: 811-818.

Kraft, Matthew A. 2018. "Interpreting Effect Sizes of Education Interventions." Working paper. Cambridge, MA: Harvard University.

Lee, Jaekyung. 2015. "College for All: Gaps Between Desirable and Actual P-12 Math Achievement Trajectories for College Readiness." *Education Researcher* 42, 1: 43-55.

Leslie, Sara-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland. 2015. "Expectations of Brilliance Underlie Gender Distributions Across Academic Disciplines." *Science* 347: 262-265.

McCallumore, Kyle Megan, and Ervin F. Sparapani. 2010. "The Importance of the Ninth Grade on High School Graduation Rates and Student Success." *Education Digest* 76, 2: 60.

McFarland, Joel, Patrick Stark, and Jiashan Cui. 2016. *Trends in High School Dropout and Completion Rates in the United States: 2013*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Mindset Scholars Network. 2015a. "Growth Mindset." Website: https://mindsetscholarsnetwork.org/learning-mindsets/growth-mindset.

Mindset Scholars Network. 2015b. "The Mindset Scholars Network." Website: https://mindsetscholarsnetwork.org.

Mueller, Claudia M., and Carol Dweck. 1998. "Praise for Intelligence Can Undermine Children's Motivation and Performance." *Journal of Personality and Social Psychology* 75, 1: 33.

National Forum on Education Statistics. 2011. *Prior-to-Secondary School Course Classification System: School Codes for the Exchange of Data (SCED)*. NFES 2011–801. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Paunesku, David, Gregory Walton, Carissa Romero, Eric Smith, David Yeager, and Carol Dweck. 2015. "Mind-Set Interventions Are a Scalable Treatment for Academic Underachievement." *Psychological Science* 26, 6: 784-793.

Somers, Marie-Andrée, William Corrin, Susan Sepanik, Terry Salinger, Jesse Levin, and Courtney Zmach. 2010. *The Enhanced Reading Opportunities Study Final Report: The Impact of Supplemental Literacy Courses for Struggling Ninth-Grade Readers*. NCEE 2010-4021. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Tipton, Elizabeth, David S. Yeager, Ronaldo Iachan, and Barbara Schneider. In press. "Designing Probability Samples to Study Treatment Effect Heterogeneity." In Paul J. Lavrakas, Michael W. Traugott, Courtney Kennedy, Allyson L. Holbrook, Edith D. de Leeuw, and Brady T. West, eds., *Experimental Methods in Survey Research: Techniques That Combine Random Sampling with Random Assignment.* New York: Wiley.

What Works Clearinghouse. 2017. *Procedures and Standards Handbook (Version 4.0)*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Yeager, David S., and Carol S. Dweck. 2012. "Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed." *Educational Psychologist* 47, 4: 302-314.

Yeager, David S., Paul Hanselman, Gregory M. Walton, Jared S. Murray, Robert Crosnoe, Chandra Muller, Elizabeth Tipton, Barbara Schneider, Chris S. Hulleman, Cintia P. Hinojosa, David Paunesku, Carissa Romero, Kate Flint, Alice Roberts, Jill Trott, Ronaldo Iachan, Jenny Buontempo, Sophia Man Yang, Carlos M. Carvalho, P. Richard Hahn, Maithreyi Gopalan, Pratik Mhatre, Ronald Ferguson, Angela L. Duckworth, and Carol S. Dweck. 2019. "A National Experiment Reveals Where a Growth Mindset Improves Achievement." *Nature* 573: 364-369.

Yeager, David, Carissa Romero, Dave Paunesku, Christopher Hulleman, Barbara Schneider, Cintia Hinojosa, Hae Yeon Lee, Joseph O'Brien. Kate Flint, Alice Roberts, Jill Trott, Daniel Greene, Gregory Walton, and Carol Dweck. 2016. "Using Design Thinking to Improve Psychological Interventions: The Case of the Growth Mindset During the Transition to High School." *Journal of Educational Psychology* 108, 3: 374.

Yeager, David S., and Gregory L. Walton. 2011. "Social-Psychological Interventions in Education: They're Not Magic." *Review of Educational Research* 81, 2: 267-301.